

09/647 875

PCT/CN 99/00046

CN 99/46

证 明

REC'D 29 APR 1999

WIPO PCT

本证明之附件是向本局提交的下列专利申请副本

申 请 日: 98 11 20

申 请 号: 98 1 25015.7

申 请 类 别: 发 明

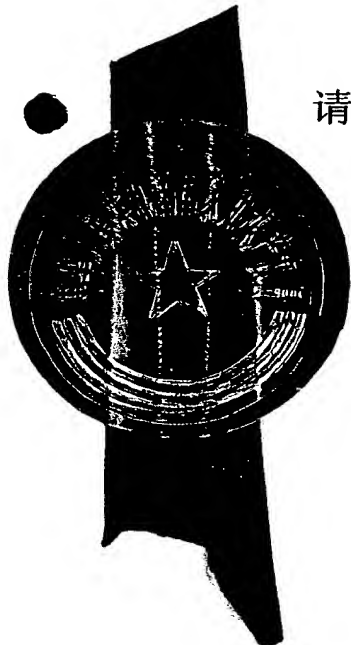
发 明 创 造 名 称: 开放式全息全选全程模板式人机互补
语言翻译方法

发明人或设计人: 刘 莎

请 人: 刘 莎

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



中 华 人 民 共 和 国

国家知识产权局局长

姜颖

99 年 04 月 09 日

权 利 要 求 书

1.一种开放式全息全选全程模板式人机互补语言翻译方法,其特征在于包括下述步骤:

- 5 a.建立一个开放的以句子为对象的多语通用的自然语言翻译人机对话模板;
 b.由自然语言翻译人机对话模板提供与原文自然语言符号项对应的多语通用的全部备选信息项及供用户扩展的空白信息项;
 c.由原文用户对自然语言翻译人机对话模板中的计算机优选信息项进行选项、补充、确认;
10 d.由翻译系统将人机互补确定的信息项转换生成译文符号,译文用户采用译文符号对原文自然语言符号项的全部备选信息项及扩展的空白信息项的求解结果进行查询。

2.根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述的开放式全息全选全程模板式人机对话语言翻译方法还包括:通过程序设计对由用户扩展的信息项进行使用频率统计,并根据使用频率统计结果确定新增通用基础概念,在翻译系统的所有语种的自然语言翻译人机对话模板中同步增添自然语言符号项及对应的信息项。

3.根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述的步骤b中,当与原文自然语言符号项对应的同语种某备选信息项中出现有空白时,用户可调用已被系统收入过的自然语言符号对其进行描述。

4.根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述的全部备选信息项按类包括自然语言符号项的概念定义、时态信息、语态信息、句法信息、上位“格”信息、词性信息、单复数信息和阴阳性信息,以及由系统提供的供人工设计的除上述信息以外的其它信息。

5.根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述的原文自然语言符号信息项是符号代码,包括原文文字和用图标或用数码表示的形式代码;原文自然语言符号包括单字、词、词组、短语和各种固定搭配的句子。

6 根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述步骤d中的转换求解结果过程是用人工标注图标、数码的方式来实现的。

7.根据权利要求1所述的开放式全息全选全程模板式人机互补语言翻译方法,其特征在于:所述步骤b中,所述的自然语言符号及其全部备选信息项是对各种自然语言常用符号语义信息项基础概念的并集。

8.根据权利要求 1 所述的开放式全息全选全程模板式人机互补语言翻译方法，其特征在于：所述步骤 b 中的多语通用备选信息项，是根据语种差异和译文精度要求差异，由系统对用户进行有选择的开放。

说明书

开放式全息全选全程模板式人机互补语言翻译方法

5 本发明涉及一种计算机翻译方法,更确切地说是涉及一种在计算机网络中适于各网络终端以不同自然语言进行信息传递交流的机器翻译方法。

10 本申请人于 1998 年 4 月 6 日向中国专利局提出名称为《全息全选全程模板式人机对话语言翻译方法》、申请号为 98101156.x 的发明专利,该方法的实现步骤是: a.建立一个以句子为对象的包括各种自然语言必要信息要素的自然语言翻译人机对话模板; b.由对话模板提供对不同自然语言进行统一通约受限后的所有备
选信息项; c.先由翻译系统对统一通约受限后的所有备选信息项进行自动优选,再由用户在全息对话模板上对优选结果进行人工调整和确认; d.由翻译系统根据
确定信息项进行译出目标语转换生成,将源语输入方的选项结果随译文提供给用
户查询。

15 全息全选全程模板式人机对话语言翻译方法作为不同自然语言语义信息传递的一种通用方式,与其它计算机自然语言翻译技术相比,在语义信息传递准确、适于多语通用方面是具有明显优势的。但是,由于人机交互信息项是经统一
通约受限过的,用户进行人机交互的方法不够多样、灵活,并且由于对系统未收
入的符号不能进行处理,因此不能满足自然语言不断发展变化的客观要求。

20 本发明的目的是设计一种开放式全息全选全程模板式人机互补语言翻译方法,作为对在先申请的 98101156.x 《全息全选全程模板式人机对话语言翻译方法》技术方案的补充,从而将全息全选全程模板式人机对话语言翻译方法扩展成为一个人机交互方式更为灵活多样、表达内容不受限制、生成译文更为自然流畅
的人机对话多语翻译方法。

25 本发明的目的是这样实现的:一种开放式全息全选全程模板式人机互补语言翻译方法,其特征在于包括下述步骤:

a.建立一个开放的以句子为对象的多语通用的自然语言翻译人机对话模板;

b.由自然语言翻译人机对话模板提供与原文自然语言符号项对应的多语通用的全部备选信息项及供用户扩展的空白信息项;

30 c.由原文用户对自然语言翻译人机对话模板中的计算机优选信息项进行选项、补充、确认;

d.由翻译系统将人机互补确定的信息项转换生成译文符号,译文用户采用译文符号对原文自然语言符号项的全部备选信息项及扩展的空白信息项的求解结果进行查询。

35 所述的开放式全息全选全程模板式人机对话语言翻译方法还包括:通过程序

设计对由用户扩展的信息项进行使用频率统计，并根据使用频率统计结果确定新增通用基础概念，在翻译系统的所有语种的自然语言翻译人机对话模板中同步增添自然语言符号项及对应的信息项。

5 所述的步骤 b 中，当与原文自然语言符号项对应的同语种某备选信息项中出现空白时，用户可调用已被系统收入过的自然语言符号对其进行描述。

所述的全部备选信息项按类包括自然语言符号项的概念定义、时态信息、语态信息、句法信息、上位“格”信息、词性信息、单复数信息和阴阳性信息，以及由系统提供的供人工设计的除上述信息以外的其它信息。

10 所述的原文自然语言符号信息项是符号代码，包括原文文字和用图标或用数码表示的形式代码；原文自然语言符号包括单字、词、词组、短语和各种固定搭配的句子。

所述步骤 d 中的转换求解结果过程是用人工标注图标、数码的方式来实现的。

15 所述步骤 b 中，所述的自然语言符号及其全部备选信息项是对各种自然语言常用符号语义信息项基础概念的并集。

所述步骤 b 中的多语通用备选信息项，是根据语种差异和译文精度要求差异，由系统对用户进行有选择的开放。

20 本发明开放式全息全选全程模板式人机对话语言翻译方法的技术特点是：由于自然语言翻译的根本目的是传递原文自然语言符号实际携带的语义信息，本发明方案通过建立开放的自然语言全息模板，为各语种中的各个自然语言符号可能携带的所有语义信息均提供了具有灵活开放特征的人机交互求解的渠道，实现全息翻译；建立必要的自然语言表达规范是完成多义符号语义信息项求解任务、保证语言自动翻译质量的重要条件，本发明用全息模板向用户提供可控制的备选信息项，从而建立起面向用户的多语种通用自然语言表达规范，通过在全息模板内

25 进行人机交互选项，来全面求解自然语言符号串实际携带的语义信息，实现全选翻译；即使已对原文自然语言符号的语义信息项进行了全面求解，在转换成了译文符号串后，仍可能新增歧义，为了保证翻译系统的原文语义信息的传递质量，本发明可根据用户需要将原文全息清单的任意信息项人机交互的最终求解结果，包括中间代码，转换成译文符号形式，以供译文用户查询，实现全程翻译。

30 本发明的开放式全息全选全程模板式人机对话语言翻译方法，在网络信息交流领域具有普遍应用的价值，在打开网络在线服务、机器翻译服务方面将会有广阔的国际市场。

下面结合一个将英文句子 “I saw a boy with a telescope near the bank”，翻译成中文的开放式全息全选全程模板式人机对话语言翻译过程进一步说明本发明的技术。

35

图 1 是全息全选全程模板式人机对话语言翻译人机交互信息处理的基本过程

框图。

图 2 是以英文原句为对象的自然语言全息模板结构示意图。

图 3 是图 2 所示模板结构的中文对照图。

5 图 4 是查询自然语言符号 “ with a telescope ” 实际携带句法信息项的过程示意图。

图 5 是用图标、数字或不同形状、颜色线条、框格作代码进行自动或人工标注的过程示意图。

图 6 是在某种语言的信息项符号空缺时对其进行语义描述的过程示意图

10 参见图 1，图中流程说明了全息全选全程模板式人机对话语言翻译人机交互信息处理的基本过程，其中中间列框 11 至 17 是翻译系统计算机的主流程，左边列框 21 至 16 示出用户的参与过程，右边列框 31 至 35 示出人机交互过程及与内部数据库、规则库间的关系，单向箭头表示人机交互流向，双向箭头表示在语言翻译过程中对数据及规则的调用过程，所标的 N 表示系统信息处理需要人机交互，所标的 Y 表示自动进入系统流程的下一个操作步骤，## - - - - ## 表示
15 此翻译系统与因特网系统的信息处理接口。其上方表示原文用户端，其下方表示译文用户端。

处理过程开始，执行步骤 11，由原文用户顺序输入待翻译的自然语言符号。

结合参见图 2 并对照图 3，在模板的序位 1 至 10 中依次填入本例中的十个自然语言符号 I saw a boy with a telescope near the bank；系统主程序的步骤 12 在可扩展的多语对应的语义信息项符号库 31 中，对各自然语言符号进行备用（语义）信息项搜索，当搜索不到时可通过步骤 21 由原文用户在模板上用系统已收入的语义符号对语义进行描述，上述过程最终生成模板中的由概念定义项、特殊属性项、时态项、语态项等构成的语义备用信息项，如果在某自然语言符号下出现概念定义信息项空白，如 bank 处出现？，则系统用户可采用系统中已提供有信息项的词汇符号对其进行语义描述，即模板中概念定义项
25 “ institution for keeping or lending money ”；系统主程序的步骤 13，根据语义信息项概率性优选规则库 32 中的规则，如模板中用黑体字指定的信息项，由计算机对列入模板中的各自然语言符号的多个语义备用信息项进行自动优选，可通过步骤 22 由原文用户对未获得确定优选的语义信息项进行选择确认；系统主程序的步骤 14，通过调用句法成份信息项自动标注模型库 33，对列入模板中的各自然语言符号的句法信息项进行自动标注，上述过程最终生成模板中的句法成份项、词性项、上位“格”项、场景项；系统主程序的步骤 15，调用句法成份信息项自动优选规则库 34，对各自然语言符号的句法成份信息项进行自动优选，
30 其间可通过步骤 24 调用句法信息项三维结构模型库 23，由原文用户在模板上对未获得唯一优选结果的句法信息项进行选择确认，如模板中用黑体字指定的信息项；系统主程序至此就可以以自定的编码形式在网络上传递所确定的上述信息
35

项。

模板中的全部信息项已扩展为自然语言符号所能携带的所有信息项，其全部备用信息项不仅包括自然语言符号的概念定义、时态信息、语态信息、句法信息、上位“格”信息、词性信息、单复数信息、阴阳性信息、而且可在开放式的模板下部扩展人工设计和标注的其它信息。

结合参见图 4，译文用户端系统主程序的步骤 16，调用译文自动转换生成规则库 35，根据多语符号与序位转换规则，将经原文用户确认的信息项求解结果自动转换成译文用户要求的自然语言译文，如图 4 中所示的中文转换生成结果“在银行附近我看见一男孩带望远镜”；系统主程序在步骤 17 将会询问用户译文是否已经无歧义，如果有歧义，译文用户可在步骤 26 通过人机交互过程确定相关信息项的查询范围，基间可调用多语对应信息项符号库 25，如译文用户为了求解“带望远镜”到底是修饰主语还是修饰宾语，如图 4 中？所示，就可直接查询该符号实际携带的句法信息项，从而确定是修饰宾语。至此翻译过程结束。

模板中的所有备选信息项均可利用图标或数字或不同颜色的线条和框格作为代码进行人工或自动标注，如图 5 中所示的用数字标注，用 1 标注“我”，用 ~ 1 标注“在银行附近”（句法信息），用 2 标注“看见”、“看见一个”（句法信息），用 3 标注“男孩”，用 ~ 3 标注“带望远镜的”（句法信息）等。

结合参见图 6，图中示出以自然语言符号“笑”为例，并集各种自然语言常用符号语义信息项形成基础概念的过程。图中仅列出其英语、汉语的基础概念并集结果，俄语、日语、法语、德语、西班牙语及其它语种可按相同方法进行语义描述。图中所示的三个主要过程是：如某种语言基础信息项符号出现空缺时，用该种语言的其它常用语义信息项符号进行语义描述，形成基础信息项，如“笑”与“laugh”；当某种语言非基础信息项空缺时，则与基础信息项符号建立模糊对应关系，形成非基础信息项，如“大笑、微笑、笑逐颜开、笑哈哈”与“burst、smile”；当某种语言其它近义信息项空缺时，则用该语言常用语义信息项符号进行语义描述，形成其它近义项，如“傻笑、咧嘴笑”与“snicker、grin”。表中的统一编码就是图 1 中的自定的编码形式，用于在网络上传递信息。

原文用户在图 1 步骤 21 中利用语义描述方法求解原文符号时，系统程序还要自动进行其使用频率的统计，在使用频率达到一定水平时，即在翻译系统收入的所有语种的自然语言符号库中同步增添新增的自然语言符号或新增信息项。如人工描述求解 bank 的使用频率达到一定水平时，系统即在法语的自然语言符号库中添加新增的符号“banque”和利用系统已收入的相应法语符号进行语义描述，并给出其它相关备选信息项。其它语言的扩展方法与其相同。

实施时，模板中的各自然语言符号项可以是单字，如“我”，也可以是词，如“银行”（或词组“国家银行”），短语，如“带望远镜的”，和各种固定搭配，如“在…附近”，当然也可以是句子（如“瞎子点灯白费蜡”）。

此外，如果用户只要求翻译基本语义，则系统开放的供人机交互的选项类则可相应减少；反之，当用户要求译文的流畅性高时，则供人机交互选项的类要相应增多。

5 本发明的方法以人机互补、多语通用、语义信息传递准确为特色，尤其适用于只懂母语的用户在计算机网络上进行广泛的多语交流。

说明书附图

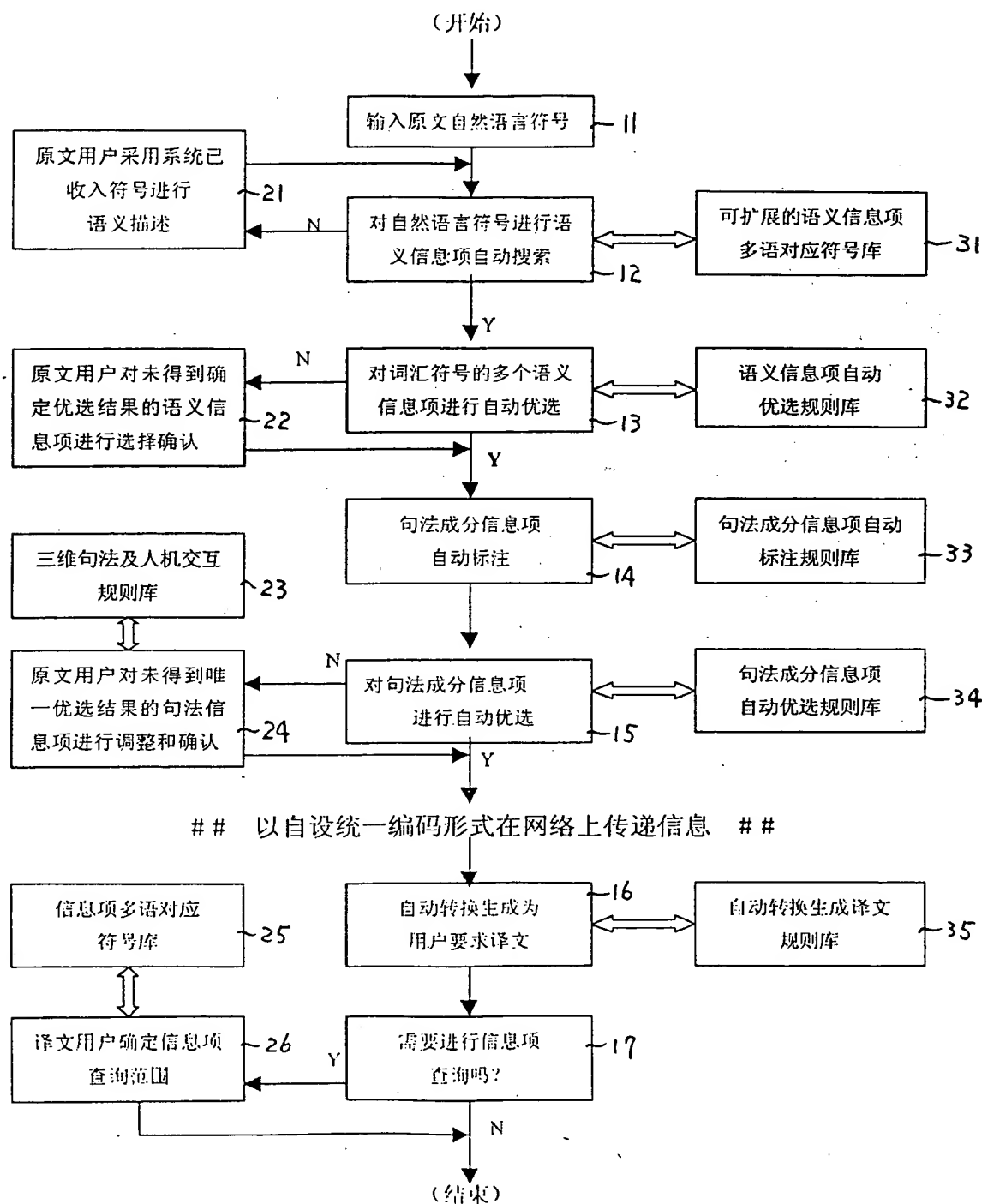


图 1

Sequence	1	2	3	4	5	6	7	8	9	10	
symbol	I	saw	a	boy	with	a	telescope	near	the	bank	.
INFORMATION ITEMS											
Definitio n of Concepts	1.the person speakin g 2. the 9th letter of the English alphabet	to use the eyes; to look at; to unders tand or recog nize	1.one; 2.the first letter of English alphabet	a young male person	1. in the presenc e of 2. beside 3.among 4.includi ng	1.one; 2.the first letter of English alphabet	a tubelike scientific instrume nt, used for making distant objects look larger or closer	1.not far from in distance , time,	1.specific person or thing. 2.a group of 3.used for comparas on	? : <i>institutio n for keeping or lending money</i>	
semantic attributes	countabl e; single		male; countabl e; single	male; counta ble; single			countable; single			countable; single	
tense		past tense									
voice		active ;passi ve									
grammar attribute	pronoun	noun; verb,	article, indefinit e	noun	preposit ion	article, indefinit e	noun verb	Preposit ion. verb. adverb.	Article, adverb	noun; verb	
part of speech	subject	predi cate	modifier of the object	object	<i>modifier of the predicate or modifier of the object?</i>			complementary of the predicate			
case(upper position)	agent	behav ior or act		Object of the behavior			tool; behavior			place behavior	
Environm ent of the sentence											

图 2

序位	1	2	3	4	5	6	7	8	9	10	11
符号	I	saw	a	boy	with	a	telescope	near	the	bank	.
	信 息 项										
概念定义	说话者本人 2、字母表的第九个字母	看见 2、理解与认识	某一 2、字母表头一个字母	男性 年轻人	带有、具有。 2、在---旁边 3、在---中间 4、包含	某一 2、字母表头一个字母	一种管状的科学仪器，用来将远处的物体放得更大更近；2、一	在时间和空间上离得不远； 2、——	特指的人或事 2、表示一类 3、用于比较	存和借贷货币的机构 2、—— -	
语义属性概念	可数 单数			阳性 可数 单数			可数 单数			可数 单数	
时态信息		过去									
语态信息		主动 被动									
语法属性	代词	名词 动词	不定冠 词	名词	介词	不定冠 词	名词 动词	副词 介词 名词	定冠词 副词	名词 动词	
句法成分	主词	谓词	宾词修 饰	宾词	谓词修饰 或宾词修饰			谓词补充			
上位语义	行为 主体	行为		行为 对象		工具 行为			处所 行为		
语言环境											

图 3

符号	在	银行	附近	我	看见	一	男孩	带	望远镜	的	
	信 息 项										
语义信息								携带 ? 附带			
句法信息	修饰主词			主词	谓词		宾词	修饰宾词			

图 4

符号	在	银行	附近	我	看见	一	男孩	带	望远镜	的	
	/		1/	1	2		3	/		3/	
语义信息								携带 ? 附带			
句法信息	~1			1	2		3	~3			

图 5

常用语义信息项 (基础概念)	统一 编码	英语符号	汉语符号	俄	日	法	德	西	---
基础信息项: 如某种语言基础信息项符号空缺, 则用该种语言其他常用语义信息项符号进行语义描述									
	01	Laugh	笑						
非基础信息项: 如某种语言非基础信息项空缺, 则与基础信息项符号建立模糊对应关系。如: 汉语的“笑逐言开”与英语的“Laugh”对应, 获得编码 01									
(程度+)	01-1-x	Burst into	大笑 -----						
(程度-)	01-2-x	smile	微笑 -----						
(成语)	01-3-x		笑逐言开 -----						
(书面语)	01-4-x								
(口语)	01-5-x		笑哈哈 -----						
(俚语)	01-6-x								
(俗语)	01-7-x								
其它近义: 如某种语言其他近义信息项空缺, 则用该语言常用语义信息项符号进行语义描述									
(褒义)	01-8-x								
(贬义)	01-9-x	snicker	傻笑 -----						
(其它近义)	01-10-x	grin	咧嘴笑						

图 6

This Page Blank (uspto)